EXHIBIT J

Vinson&Elkins

Alden L. Atkins aatkins@velaw.com
Tel +1.202.639.6613 **Fax +**1.202.879.8813

March 22, 2021

<u>Via Email</u>

Sarah R. LaFreniere
Hausfeld LLP
888 16th Street N.W.
Suite 300
Washington, DC 20006

Re: Defendants' methodologies for technology assisted review ("TAR")
       In re Diisocyanates Antitrust Litigation, MDL No. 2862

Dear Counsel:

I write to follow up regarding our meet and confer on March 19, 2021 regarding Domestic Defendants' disclosed TAR methodologies and to respond to your letter dated March 18, 2021.

Though the parties have spent hours meeting and conferring and have exchanged multiple lengthy letters on this issue, it has become evident that the Plaintiffs do not have a principled argument against Defendants' proposed methodology, nor the statistics that support it. Rather, your March 18 letter, reinforced by your comments during the March 19 meet and confer, highlight the ultimate issue—your concern is not with the validity of Defendants' proposed TAR methodology, but rather, it is that Defendant's proposed methodology is not *Plaintiffs'* proposed methodology. As you stated in your letter, Plaintiffs "request that Defendants adopt Plaintiffs' proposed validation proposal." Defendants have proposed a straight-forward, industry standard approach to implementing and validating TAR. As the Sedona Principles provide, "Responding parties are best situated to evaluate the procedures, methodologies, and technologies appropriate for preserving and producing their own electronically stored information."[1] The Court's Stipulated Order Re: Discovery of Electronically Stored Information ("ESI") Protocol requires simply that, "If a party decides to use a computer/technology-assisted review ("TAR") for identifying potentially responsive documents, it shall inform the other parties and, if requested, shall discuss methods and procedures for implementing it." (ECF No. 313) We have done that, and absent a showing from Plaintiffs that there is a deficiency with Defendants' proposals, which again, are

---

[1] The Sedona Principles, Third Edition: Best Practices, Recommendations & Principles for Addressing Electronic Document Production, 19 Sedona Conf. J. 1 (2018), Principle 6.

V&E

commonplace, there is no basis for Plaintiffs' seeking to unilaterally impose their preferred TAR procedures on Defendants.

Defendants responded to what they understood to be a key concern for Plaintiffs—that Defendants had proposed initiating a review using an estimation/richness sample of a size derived based on a 95% confidence level and a 5% margin of error—by agreeing to employ a margin of error of 2%. Though Plaintiffs "appreciated" Defendants' movement, the position in your latest letter has evolved again. Now you say you "do not believe a richness or estimation sample is necessary." Fine—and Plaintiffs may be able to design a TAR methodology that operates without one—but it certainly would not be industry standard, nor would it necessarily offer anything that Defendants' methodology does not.

Richness/Estimation Sample

To recap, and as expressed in our last letter, we had previously understood that Plaintiffs' concern was principally regarding Defendants' proposed estimation sample methodology—as you asserted in your March 4, 2021 letter that "Defendants' statistical measure for deriving the richness sample will yield a large margin of error" and devoted more than a page of text to attacking the statistical measure. In response, Defendants explained in our March 12 meet and confer, and reiterated in our March 16 letter, why Plaintiffs' math was wrong, but how, nevertheless, to compromise Defendants agreed to sample to a statistical confidence level of 95%, with a margin of error of +/- 2%.

Specifically, in our previous letter we explained the following:

> Plaintiffs' March 4 letter presented a hypothetical in which a +/-5% margin of error would mean that the actual number of responsive documents in a population with a richness of 10% would be between 50,000 and 150,000. In reality, the margin of error depends on the richness point estimate; that is, mathematically, the margin of error reaches its highest level only when, point estimate of richness is 50%. As we explained, with a sample designed to achieve a +/- 2% margin of error would actually be 1.12% at a 10% richness. This means that one would be 95% confident that the actual number of responsive documents in the review set is between 88,000 and 112,000, a reasonable and narrow margin.

Instead of responding to this, Plaintiffs present a new hypothetical but commit the same math error. Now Plaintiffs ask us to consider a collection of one million documents with a 20% prevalence. Plaintiffs' hypothetical again overlooks the fact that the actual margin of error is dependent on the richness point

V&E

estimate; that is, the margin of error reaches its highest level only when the point estimate of richness is 50%. Plaintiffs' hypothetical presents a richness sample where the prevalence estimate is 18%, but the actual prevalence is 20%. Such a circumstance would fall outside the +/- 1.6% margin of error of a sample where 20% of the documents are responsive. Based on a confidence level of 95%, this means Plaintiffs' hypothetical scenario can happen less than 5% of the time. That is, though it is theoretically possible for the sample to be at the extreme end of the margin of error, there is a probability curve, with the greatest probabilities clustered around 200,000. There could be a sample that estimates 184,000 (i.e., at the far low end of the correct margin of error) when the real number is 200,000, but this is would occur very rarely, only 5% of the time. That is a statistically acceptable risk to take as a measure of reasonableness.

In addition, your new hypothetical is unrealistic and therefore is unhelpful for our discussion. You have assumed a richness of 20%, yet published research reports that the responsiveness rate is almost always between 1% and 5%. At a more realistic responsiveness rate of 3%, the margin of error would actually be +/- 0.68%. In your hypothetical collection of 1 million documents, that means we could be 95% confident that the true number of responsive documents would be 23,200 to 36,800, which by any measure is a reasonable and acceptable range.

So again, though Plaintiffs now say that they do not feel that a Richness/Estimation Sample is necessary, they do not provide any rationale for why it is not an appropriate metric by which to decide when to pause the review and start the validation process. Again, as stated in our March 16 letter, all Defendants propose to review documents until each Defendant reasonably believes it has achieved a targeted recall rate of approximately 70%. The Richness/Estimation Sample is simply a guide for when to start the validation testing. Defendants then would use an elusion sample, which reflects a randomly drawn sample with 95% confidence and 2% margin of error from the unreviewed documents in the TAR review set, in combination with the number of responsive documents already found, to determine with a high degree of statistical confidence that the targeted recall rate was actually achieved against the entire corpus of documents in the TAR review set. As stated, each Defendant will resume reviewing if the targeted Recall rate has not been achieved.

Stopping/Validation Criteria

We will not repeat all of our objections to the Plaintiffs' arbitrary proposed validation process, except to note the following.

Claim that Defendants' proposed stopping/validation criteria is insufficient. Plaintiffs state that "Using Defendants' estimate of recall (rather than true recall), the possibility of missing a substantial number of relevant documents is not infinitesimal at all, and academic literature has demonstrated that

V&E

estimates of achieving target recall are often woefully inaccurate." Again, Plaintiffs overlook that validating using statistical measures is a commonly accepted methodology.[2] In our previous letter, we excerpted from the 2019 EDRM Technology Assisted Review (TAR) Guidelines ("TAR Guidelines"), which represent a consensus document based on the input of "[m]ore than 50 volunteer judges, practitioners, and e-discovery experts."[3] To reiterate, the TAR Guidelines recognize "two primary approaches to estimating the extent to which TAR has found relevant documents"—Defendants' proposed methodologies incorporate one or both of these.

Inclusion of the "Delta Set" in the validation of procedures. You have repeatedly referenced the _Broiler Chickens_ TAR methodology,[4] and your latest letter states that Plaintiffs' validation procedure, described in Appendix A of your February 19, 2021 letter, "was developed from the Broiler Chickens Case." You further state that "Plaintiffs' proposal simply breaks down the document collection into four subcollections, to determine if there is one particular category that sampling is more prone to error. Such _stratified sampling_ is a well-established statistical method to improve accuracy."

Plaintiffs' proposed procedure may have been "developed from" the _Broiler Chickens_ methodology, but it departs from those procedures in at least two critical respects. First, search terms were used to significantly cull nonresponsive documents from the set searched with TAR in _Broiler Chickens_, while Plaintiffs here continue to insist on woefully overbroad search terms. Second, the _Broiler Chickens_ methodology did not perform validation sampling of the documents removed using search terms. In contrast, Plaintiffs' proposal here includes validating TAR in part based on a review of documents "excluded from TAR"—Plaintiffs' so-called "Delta Set." As even the Wikipedia page that you cite recognizes, "stratified sampling is a method of sampling from a population which can be partitioned into subpopulations. . . . Stratification is the process of dividing members of the population into _homogeneous subgroups_ before sampling."[5] A set of documents that was generated using search terms designed to target responsive documents can hardly be said to be homogenous with a different set of documents excluded from the TAR review set because those documents hit on none of the extremely broad search terms.

---

[2] _See, e.g.,_ Lea Malani Bays & Lexi J. Hazam, Technology-Assisted Review: Advice for Requesting Parties, Oct./Nov. 2016, https://www.lieffcabraser.com/pdf/20161000_LJH_PracticalLaw.pdf, at 7 ("Any method used to calculate recall should involve a large enough sample so that the margin of error is at an acceptable level. For example, a 95% confidence level and a 2% margin of error is usually considered an effective sample size.")

[3] EDRM, Guidelines Aim to Define and Demystify Technology-Assisted Review and to Encourage Widespread Adoption of TAR Processes, Feb. 7, 2019, https://edrm.net/2019/02/edrm-releases-tar-guidelines/.

[4] Order Regarding Search Methodology for Electronically Stored Information, _In re: Broiler Chicken Antitrust Litig.,_ No. 1:16-cv-08637 (N.D. Ill. Jan. 3, 2018).

[5] Wikipedia, "Stratified Sampling," https://en.wikipedia.org/wiki/Stratified_sampling (emphasis added).

V&E

While no defensible statistically-valid conclusions can be drawn based on such a procedure,[6] what it does allow for is needless second-guessing and 11th-hour questioning of the TAR review process. Instead, Defendants' proposed procedures rest on objective, statistically significant, commonly accepted measures.

Claim that Plaintiffs' Proposal Imposes no Additional Burden. Plaintiffs write that their proposed stopping criteria and validation proposals "would result in sampling of almost the same number of documents (approximately 4,800 in Defendants' proposal and 5,000 in Plaintiffs')." Merely reviewing that number of documents is not all that is contemplated by the validation procedures. This aspect of Plaintiffs' proposal compounds the issues with the Delta Set noted above, as it may require additional review to achieve targeted recall based on the responsiveness characteristics of an unrelated, dissimilar set of documents—a measure which, again, has no statistical significance.

Stopping Criteria. As we have discussed, another concern is that Plaintiffs' proposed stopping criteria require that "none of the responsive documents is novel and/or more than marginally relevant." As we have said, this is a subjective criteria which would invite endless disputes about whether any document in the last set was "novel" or "more than marginally relevant." While Plaintiffs express confidence that the parties could agree on objective standards for this inherently subjective criteria, we have asked twice for Plaintiffs to provide an objective, workable definition of "novel" and "no more than marginally relevant" and we have not received an answer. You said that you would consider providing a definition. This raises further proportionality concerns as we understand that new elusion tests would need to be run and additional documents reviewed, if an unspecified "substantial number" of "novel"/"more than marginally relevant" documents are identified.

Therefore, Defendants do not accept Plaintiffs' proposal. It is our hope that Plaintiffs will recognize Defendants' prerogative to design and implement their own TAR methodologies. Defendants have done so based on objective criteria that are grounded in commonly accepted statistical metrics and based on the experience of our several e-discovery vendors.

---

[6] During our meet and confer, we asked the basis for the number of documents Plaintiffs have proposed for each of the four sets in its validation procedure. As best we can tell, they were not chosen to achieve any defined level of statistical confidence. You did not know, but you said you would consult with your team and let us know. It is not clear why Plaintiffs have taken issue with Defendants' proposed sampling methodology when Plaintiffs themselves employ some form of sampling in the validation stage (though the basis for those sample sizes is unclear) as well as in Plaintiffs' proposed elusion test (though with a 5% margin of error, as opposed to Defendants' 2%).

V&E

Sincerely,

/s/ Alden L. Atkins

Alden L. Atkins

Cc: Megan E. Jones
        Jason S. Hartley
        William Pietragallo, II